Exhibit F

The MEDTAP Institute at
UBC

A2-3759

Report Prepared for:

Mundipharma Research GmbH &Co. KG
Limburg, Germany and its independent Associates

Validation of the Bowel Function Index

Prepared by:

Anne M. Rentz, MSPH
Lori Frank, PhD

7101 Wisconsin Avenue
Suite 600
Bethesda, MD 20814

PHONE: 1-301-654-9729
FAX: 1-301-654-9864

www.UnitedBiosource.com

June 15, 2005
Revised: July 12, 2005

TABLE OF CONTENTS

1.0 BACKGROUND AND STUDY OBJECTIVE 1

2.0 STUDY METHODS 1

2.1 Patient-Reported Outcomes 2
2.1.1 Bowel Function Index 2
2.1.2 Global Tolerability 3
2.1.3 Additional Data Collected 3
2.2 Analysis 3
2.3 Analysis Plan 3
2.3.1 Reliability 4
2.3.2 Validity 5
2.3.3 Responsiveness and Clinical Significance 6

3.0 RESULTS 7

3.1 Sample 7
3.2 Item Performance 7
3.3 Reliability 8
3.3.1 Internal Consistency Reliability 8
3.3.2 Reproducibility (Test-Retest Reliability) 9
3.4 Validity 9
3.4.1 Concurrent Validity 9
3.4.2 Discriminant Validity 10
3.5 Responsiveness 10
3.5.1 Effect Size 10
3.5.2 Standard Error of Measurement (SEM) 11
3.5.3 One half SD 11

4.0 DISCUSSION 11

5.0 OVERALL CONCLUSIONS 13

6.0 REFERENCES 14

7.0 TABLES 15

Table 1. Demographic Characteristics at Visit 2 16

Table 2. Clinical Characteristics at Visit 3[a] 16

Table 3. Diarrhea as Reason for Discontinuation 16

Table 4. Constipation Item Analysis, Visit 3 and Visit 5 17

Table 5. Inter-Item Correlations and Internal Consistency Reliability: Constipation Items, Visit 2 18

Table 6a. Reproducibility: Constipation Items, Visit 5 and Visit 6 Data for Patients Randomized to Naloxone Placebo Group 19

Table 6b. Reproducibility: Constipation Items, Visit 5 and Visit 6 Data for Stable[a] Patients Randomized to Naloxone Placebo Group 19

Table 7. Concurrent Validity: Correlation[a] between Constipation Items and Clinical Data, Visit 5 20

Table 8. Concurrent Validity: Constipation Item Score Comparisons, Data from Visit 5[a] 20

Table 9. Discriminant Validity: Constipation Severity based on Stool Consistency Rating in Diary (Analysis of Variance[a]), Visit 5 21

Table 10. Responsiveness of Constipation Items, Visit 3 and Visit 5, by Treatment Group 22

1.0 BACKGROUND AND STUDY OBJECTIVE

Opioid analgesics are effective at pain relief, but the mechanism of action leads to slowed bowel motility and constipation. Mundipharma has completed a Phase II clinical trial of a combination oxycodone/naloxone product, OXN 2401, and collected data on analgesia efficacy and side effects, including constipation. The primary objective of that trial was to determine whether an oxycodone/naloxone combination leads to a comparable analgesia with a decrease in constipation in patients with severe chronic pain of tumor and non-tumor origin compared to oxycodone while maintaining an acceptable tolerability profile. Establishing that patient-reported items on constipation exhibit favorable psychometric properties is fundamental to continued use of these items in clinical trials and interpretation of the data they generate (Revicki et al. 2000; Leidy et al. 1999).

Mundipharma reviewed the literature to locate an instrument for assessment of constipation from the patient's point of view. There were no available patient-reported instruments appropriate for a sample using opioid analgesics. Based on the literature, the Rome criteria (Drossman et al. 2000), and clinical judgment Mundipharma created the Bowel Function Index, a three-item questionnaire to measure constipation from the patient's perspective.

The objectives of this study are (1) to evaluate the psychometric properties of the Bowel Function Index (BFI), three patient-reported questions on constipation used in Mundipharma's recent Phase II clinical trial; and (2) to evaluate the responsiveness and clinical significance of the BFI.

2.0 STUDY METHODS

The psychometric analyses were performed as secondary analyses on data collected during Mundipharma's OXN 2401 trial, a multicenter, prospective, controlled, randomized double-blind 4 parallel group Phase 2 study of oral controlled release (CR) oxycodone, oral CR naloxone and corresponding naloxone placebo. Patients had severe chronic pain of tumor or non-tumor origin that required opioid treatment, accompanied by constipation. Patients with an analgesic treatment stabilized at 40mg, 60mg, or 80mg per day of CR oxycodone were randomized to 10mg, 20mg, or 40mg per day of CR naloxone or to CR naloxone placebo.

Patients answered the BFI at each visit. See Table 1 for the schedule of assessments.

Table 1. Summary of Visits and Questionnaire Completion by Participants[1]

Screening	Titration/run-in with oxycodone only		Maintenance Phase Double-blind treatment	Maintenance Phase	Follow-up Phase Oxycodone only
Visit 1 Week -1	Visit 2 Week 0 Baseline	Visit 3 Visit 2 + 1-3 weeks End of titration/run in Randomization	Visit 4 Visit 3 + 1 week	Visit 5 Visit 3 + 4 weeks End of naloxone treatment	Visit 6 Visit 5 + 2 weeks End of follow up
BFI	BFI	BFI	BFI	BFI	BFI

[1] From CSR OXN 2401 December 10, 2003.

Data were obtained from Purdue April 15, 2005 as cleaned SAS-ready datasets. All analyses were completed using SAS version 8.2.

2.1 Patient-Reported Outcomes

2.1.1 Bowel Function Index

Three questions were used in the clinical trial to assess constipation from the patient's perspective, rated on a numerical analogue scale (NAS) from 0 (good) to 100 (bad), referred to as the Bowel Function Index (BFI):

1. Ease of defecation (NAS) during the last 7 days according to patient assessment (0 = easy/no difficulty; 100 = severe difficulty)
2. Feeling of incomplete bowel evacuation (NAS) during the last 7 days according to patient assessment (0 = not at all; 100 = very strong)
3. Personal judgment of patient (NAS) regarding constipation during last 7 days (0 = not at all; 100 = very strong)

The three constipation questions are averaged to get a summary score (total score range: 0 – 100). Additionally, each question is used on its own (item score range: 0 – 100). The mean of all three items was used as a primary endpoint; individual items were used as secondary endpoints.

2.1.2 *Global Tolerability*

Global assessment of tolerability was asked of the patient and investigator at Visit 5 (end of maintenance phase week 4). It was answered on a 7-point Likert-type scale (Very Good to Very Poor). Correlations between the investigator and patient global assessment of tolerability were examined. The correlation coefficient was 0.87; as a result only the analyses based on the patients' global assessment are presented in the concurrent validity analyses.

2.1.3 *Additional Data Collected*

Patients were asked to complete a daily diary, beginning with the baseline visit (Visit 2). Relevant data from the diary were used in these validation analyses:

- Stool frequency: number of bowel evacuations per day.
- Stool consistency: median patient rating on a 4-point response scale (hard, solid, semi-solid, diarrheal).
- Number of days of laxative intake.
- Reason for discontinuation: subjects discontinuing due to diarrhea were selected for subset analyses.

2.2 Analysis

The evaluable population for these analyses were all randomized subjects who received trial medication and completed any of the three BFI constipation questions. Sociodemographic data from Visit 2 were used to reflect characteristics of the sample at baseline. Visit 3 (end of titration/run in) data were used in analyses for the pre-treatment baseline and Visit 5 (end of double-blind treatment maintenance phase) data were used as the endpoint data. Visit 5 to Visit 6 was used as the retest interval given that stability was assumed to be greatest from end of maintenance phase to end of follow-up phase. No center effects were evaluated and there was no adjustment for multiple comparisons.

2.3 Analysis Plan

Descriptive statistics are presented on sociodemographic variables from Visit 2 (age, gender, race; Table 1). Mean and median for the following clinical variables are presented based on

Visit 3 data (selected due to availability of diary data in week preceding Visit 3): daily pain intensity, stool frequency, stool consistency, and number of days of laxative intake (Table 2). Number of subjects indicating "diarrhea" as reason for discontinuation is presented in Table 3.

To assist with evaluation of item performance, descriptive statistics are presented for each constipation item, using Visit 3 and Visit 5 data: mean, standard deviation (SD), range, median, % at floor value, % at ceiling value, and % missing (Table 4). Comparison was made between each of these values at Visit 3 and at Visit 5 (endpoint) as a descriptive means of determining if item performance changes over time.

2.3.1 Reliability

Internal consistency reliability was evaluated based on Cronbach's alpha using Visit 2 data (Hays et al. 1998; Nunnally & Bernstein 1994) (Table 5). Values above 0.70 are generally considered indicative of good internal reliability. Alpha with item deleted was also examined for each item; inflation of the value more than 10% above total score (3-item mean) alpha upon item removal is considered indicative of a potentially internally inconsistent item. Inter-item correlations were also examined to evaluate internal consistency of the items, with correlations of 0.40 or less considered low inter-item correlation (Cohen 1988).

Test-retest reliability was examined in the subgroup of patients randomized to the naloxone placebo group, using Visit 5 to Visit 6 as the retest interval. Intraclass correlation coefficients (ICC), Pearson's correlations, and change in scores (via t-test to determine statistical significance of change) were calculated between Visit 5 and Visit 6 to evaluate test-retest reliability (Table 6A). These analyses were also run on the subgroup of naloxone placebo subjects who had no change in stool frequency from Visit 5 to Visit 6. The ICC quantifies strength of correlation but incorporates information on slope and intercept to address the limitations of the product-moment correlation for detecting systematic change (Deyo et al. 1991). The more stability in the measure, the higher the correlation coefficient and the ICC are expected to be.

2.3.2 *Validity*

Validity of an instrument refers to the extent to which an instrument measures the construct it is intended to measure (Hays et al. 1998; Nunnally & Bernstein 1994). *Concurrent validity* refers to the relationship of the instrument to other similar evaluations. To examine concurrent validity, the relationship between the three constipation questions and total score and selected clinical characteristics (stool frequency and consistency, number of days of laxative intake, and global assessment of tolerability) were analyzed using Spearman's rank correlations (Table 7). Visit 5 (endpoint) data were used to allow evaluation based on outcomes best measured at endpoint (e.g., number of days of laxative intake). Concurrent validity is supported when the total and item scores of the three constipation items are substantially correlated (>0.40) with items or scales measuring similar concepts (Cohen 1988). Conversely, items or scales measuring different concepts should have smaller correlations (<0.40). The following are the concurrent validity hypotheses:

1. Null hypothesis: No positive correlation between the BFI items and the following variables: laxative intake (data from 7 days following collection of constipation item data) and patient rating of tolerability. Alternative hypothesis: a statistically significant positive correlation exists between the BFI items and the stated variables (Table 7).
2. Null hypothesis: No inverse correlation between the BFI items and stool frequency or stool consistency (data from 7 days prior to collection of constipation item data). Alternative hypothesis: statistically significant negative correlation between BFI and stool frequency (Table 7).
3. Null hypothesis: No BFI score difference between patients who remain in the study to Visit 5 and those who discontinue due to a side effect of diarrhea. Alternative hypothesis: patients who discontinue due to diarrhea will have lower BFI scores than patients who remain in the study (Table 8).

4. Null hypothesis: No BFI score difference between patients who prefer the maintenance therapy phase and have better scores on the constipation items than patients who prefer the titration therapy phase. Alternative hypothesis: BFI scores will be higher for patients who prefer the maintenance therapy phase relative to patients who prefer the titration therapy phase (Table 8).

Discriminant validity is the extent to which scores from an instrument are distinguishable from groups of subjects that differ by a key indicator, usually clinical in nature. To evaluate discriminant validity, analysis of variance (ANOVA) models were used to compare the constipation item and total scores by severity level based on Visit 5 (endpoint) data (Table 9). Patients were stratified into three severity levels (mild, moderate, severe) based on stool consistency, using diary data (mean from 7 days prior to Study Visit 5), with mild = loose, moderate = soft or normal, and severe = hard. The null hypothesis is no difference between patients classified as mild and those classified as severe. The alternative hypothesis is that BFI scores for patients classified as mild will be statistically significantly lower than patients classified as severe.

2.3.3 *Responsiveness and Clinical Significance*

Responsiveness to true change over time was examined using effect size. Standard error of measurement (SEM) was used as one basis to quantify clinical significance of specific BFI point score differences from the perspective of the individual patient (Wyrwich et al. 1999) and examination of one half of one standard deviation (Norman et al. 2003) was used as another means of determining clinical significance.

Effect size is a quantitative measure of change in score, and provides a means of standardizing the quantification for comparison between groups and a means of supplementing statistical testing to provide a more comprehensive view of item or instrument performance for health status measurement (Kazis et al. 1989). Effect size 1 is defined as the mean difference pre- to post- treatment (Visit 3 to Visit 5) divided by the standard deviation of all subjects at pretreatment (Visit 3). The second estimate of effect size, effect size 2, also called Guyatt's responsiveness statistic, is a variation of the above effect size using the same numerator but limiting the denominator to the standard deviation of score changes among stable patients only

(mean score change / standard deviation of score changes among stable patients) (Kazis et al. 1989; Guyatt et al. 1987). Stable subjects are defined as those who have less than or equal to 25% decrease on the judgment of constipation item from Visit 3 to Visit 5. Effect sizes are calculated by treatment group and are one means of benchmarking less important and more important score change magnitudes (Kazis et al. 1989).

3.0 RESULTS

3.1 Sample

Table 1 contains the demographic characteristics of the patient sample included in the analysis. There were 202 subjects in the evaluable sample. The mean age was 56.3 with a range of 27 to 86 years of age. Almost two-thirds of the sample was comprised of women (63%) and all were Caucasian (100%).

Table 2 contains information on the clinical characteristics of the sample at baseline, calculated as a weekly average based on the week prior to Visit 3. The average daily pain intensity was 38 with a range of 0 to 81. The daily average stool frequency was 1 with a range of 0.1 to 4. Subjects reported that the average weekly consistency was 2.4 on a 1 to 4 scale. Subjects used laxatives an average of 6 times a week in the week before Visit 3.

3.2 Item Performance

Item performance data are presented in Table 4. There were 202 patients for whom BFI data were available at Visit 3. There were low rates of responses at floor or ceiling values. High rates at floor (best response) would limit sensitivity to detect changes over time. High rates at ceiling (worst response) might indicate poor measurement of constipation severity. Neither held true in this sample. Of the items, item 2 showed the highest rate of floor effect, indicating that the symptom of incomplete evacuation is less severe among this sample than the ease of defecation or the overall rating of constipation.

There were 169 patients for whom BFI data were available at Visit 5. More patients rated their constipation symptoms at the best possible level on the scale following treatment than before, as expected for a treatment intended to improve bowel motility. Correspondingly, means and

median BFI values were lower following treatment than before, as expected following treatment. Fewer than 27% of the sample reported values at floor indicating adequate score distribution for the BFI items. Patient responses spanned the full range of possible values from 0 to 100. Lack of missing data suggests that completion of these items was not difficult or confusing for patients.

Inter-item correlations are presented in Table 5. Items 1 and 3 had the highest correlation, 0.86. Item 2 correlated 0.59 and 0.60 with items 1 and 3 respectively. These correlation results suggest that information obtained from items 1 and 3 is more highly related than information obtained from item 2 relative to item 1 or 3.

Correlations of the items with total score are high, an expected result in a measure with so few items. Item-total correlation results are consistent with inter-item correlation results, showing the correlation of item 2 to the total BFI score to be slightly below the correlation of the other items to total BFI score. All correlation coefficients are well above the accepted threshold of 0.70, indicating strong association, as expected.

3.3 Reliability

3.3.1 Internal Consistency Reliability

Table 5 presents Cronbach's alpha for all items and total score. Internal consistency is very good, with an alpha exceeding 0.70. Deletion of item 2 increases the alpha slightly, suggesting that the relation of item 2 to the other BFI items is smaller than the relation of items 1 and 3 to the other BFI items. These results are consistent with the smaller magnitude of correlation between item 2 and the other items, relative to the item 1 and item 3 correlations. Deletion of item 1 decreases the alpha, as does deletion of item 3, supportive evidence for the value of items 1 and 3 to the cohesiveness of the BFI. When measuring a single construct it is desirable to have high internal consistency of items, although items that are extremely highly correlated may be conveying redundant information. In an extremely brief 3 item inventory like the BFI, item redundancy is not a concern.

3.3.2 Reproducibility (Test-Retest Reliability)

Reproducibility was examined in two different ways: first using data from patients randomized to the naloxone placebo group from Visit 5 to 6, and second using data from only those naloxone placebo patients who also reported no change in stool frequency. Results are presented in Tables 6a and 6b. The second method is more restrictive and therefore results were expected to be the same or better than for the first method. However, the small sample size that resulted from the selection criterion (N=12) severely limits conclusions that can be drawn from Table 6B.

Based on the 44 patients randomized to naloxone placebo, mean differences were <5 points and not statistically significant suggesting minimal change from Visit 5 to Visit 6. This supports expectations of stability in BFI scores during a period of limited expected change in constipation symptoms.

The ICC was examined to supplement the information obtained from the Pearson's product-moment correlation. Constipation symptoms fluctuate and the correlation coefficients between 0.53 and 0.72 are in the moderate to high range.

3.4 Validity

3.4.1 Concurrent Validity

Correlations between the BFI and selected clinical characteristics (stool frequency and consistency, number of days of laxative intake, and global assessment of tolerability) are presented in Table 7. Correlations between the BFI item and total scores and stool frequency and stool consistency were inverse, statistically significant, and in the low to moderately range (-0.23 to -0.29). The BFI item and total scores were also statistically significantly correlated with the patient rating of tolerability (range: 0.23 to 0.34). Correlations between the BFI total score and two of the three BFI items and number of days on laxative were greater than 0.40 and statistically significant. Item 2 had a lower correlation (0.33) suggesting that the BFI total score and items 1 and 3 are more highly related to laxative intake than is item 2.

Mean scores are statistically significantly lower for patients preferring maintenance phase relative to titration phase, as expected (Table 8). Maintenance phase preference suggests that the

medication had a favorable effect on constipation, and accordingly the scores for those patients preferring maintenance phase are lower than for the other patients.

The comparison between subjects who completed the study and those who discontinued due to diarrhea was not included because data for only one discontinued subject were available.

3.4.2 Discriminant Validity

Patients were divided into three groups based on response to the stool consistency question at Visit 5. Those who reported hard stools were classified as severe, those who reported normal or soft were classified as moderate and those who reported loose stools were classified as mild (Table 9). BFI scores differed between patients classified as moderate and those classified as severe. BFI scores also differed between patients classified as mild and those classified as severe. Differences between mild and moderate were not statistically significant. These results support the discriminant validity of the BFI, indicating that BFI score magnitudes correspond to level of constipation severity based on stool consistency as a criterion for severity.

3.5 Responsiveness

3.5.1 Effect Size

Effect size was calculated based on the mean difference pre- to post-treatment (Visit 3 to Visit 5); effect size 1 used the standard deviation of all subjects at pretreatment (Visit 3) as the denominator and effect size 2 used the standard deviation of score changes among stable patients only as the denominator (Kazis et al. 1989; Guyatt et al. 1987). Stable subjects are defined as those who have less than or equal to 25% decrease on the judgment of constipation item from Visit 3 to Visit 5. The results are shown in Table 10. Using Cohen's (1988) effect size criteria of 0.2 representing small change, 0.5 representing moderate change, and 0.8 representing large changes, the effect sizes are of the expected magnitude. Further, the effect sizes increase by naloxone dose in the expected direction, with largest effect sizes for patients at the highest dose. Effect size is lowest for item 2 relative to the other 2 items, suggesting limited responsiveness to change over time as measured by this item. The effect sizes calculated using the two different methods are substantially similar, supportive evidence in favor of the magnitude of effect sizes observed in this sample.

3.5.2 *Standard Error of Measurement (SEM)*

The SEM was calculated as one means of establishing ranges for clinically significant score change on the BFI (e.g., Norman et al. 2003; Wyrwich et al. 1999). The SEM value is shown in Table 10. For all subjects at Visit 3, the SEM value is 9.01, suggesting score changes of 9 points or greater may be clinically significant from the perspective of the individual patient.

3.5.3 *One half SD*

An SEM value close to one half SD provides converging evidence of clinical importance. Visit 3 SD for the BFI total score is 22.6 (see Table 4); one half of the SD is 11.3. Given suggestions that one-half SD may bound a lower limit for clinical significance, this suggests that BFI total score differences below 11 points may be below the threshold for clinical significance. Further evaluation would be required to determine the clinical significance of score differences between 9 as suggested by the SEM and 11 points as suggested by one half SD.

4.0 DISCUSSION

The BFI is a patient-based rating of constipation. The analyses reported here indicate that the BFI meets criteria for basic psychometric performance. The item performance data suggest that the items are feasible to administer and in this sample did not result in substantial floor or ceiling effects. Therefore, the ability of these items to measure the condition of interest and to detect true changes in that condition are not limited by the response scale.

The items of the BFI relate to one another as expected. Items 1 and 3 showed substantial overlap based on correlations; conceptually the content is distinct mitigating any concerns regarding redundancy. Content of item 2 does not overlap with items 1 and 3 as much as they overlap with each other. Incomplete bowel evacuation is likely distinct from ease of defecation in terms of symptom experience; less overlap with judgment regarding constipation suggests that the judgment is based more on defecation ease than on feeling of incomplete evacuation.

The BFI items are internally consistent, suggesting they all measure the same or substantially related constructs. Item 2 may have a slightly different relationship to that construct than items 1 and 3 but performance of item 2 supports its inclusion in the BFI. Results for item 2 internal

consistency support the results found with inter-item correlations and suggest that ease of defecation contributes more to overall judgment regarding constipation than does the feeling of incomplete evacuation. Contribution of all items is above accepted thresholds however and all items contribute meaningfully to the total BFI.

The BFI demonstrated reproducibility over time. The magnitude of correlations was in the moderate range. Interpretation of the reproducibility data must allow for the possibility of some true change occurring in subjects during the retest interval. Specifically, the lower correlation coefficient for item 1 relative to the other items suggests that the patient experience of ease of defecation did vary across the timepoints examined, a conclusion in line with the clinical course of symptoms.

The relationships between BFI scores and related pâtient reports regarding stool frequency and stool consistency were in the expected direction and all correlation coefficients met criteria for statistical significance. The number of days on laxative related directly to BFI items (e.g., the more days on laxative in subsequent week, the more trouble with constipation), consistent with the expected patient symptom experience subsequent to laxative use.

The global rating of tolerability assessment showed low to low-moderate relationship to BFI score. While constipation symptoms are just one part of a tolerability profile, these results emphasize that they are a substantial part of that profile. Relatedly, patients who expressed a preference for ongoing maintenance therapy showed better constipation resolution than patients who expressed a preference for the titration phase, as measured by BFI score. Patients expressing a preference for the maintenance phase should be those patients who achieved pain control with an acceptable side effect profile, relative to the pain control/side effect profile experienced during the titration phase. These patients would be expected to have fewer constipation symptoms or symptoms of lesser severity than other patients and the BFI data support this explanation, lending credence to the validity of the BFI as a measure of constipation.

When patients were divided into constipation severity groupings on the basis of stool consistency, BFI scores again show the expected patterns, with those patients with the most severe constipation having higher BFI scores than the other patients. While consistency has limitations as a proxy for defecation ease, incomplete evacuation and judgment regarding

constipation, it makes clinical sense to expect that it is a reasonable proxy for the purposes of demonstrating discriminant validity of the BFI.

The BFI showed responsiveness to expected constipation changes over time, and the effect sizes examined increased in a dose-response fashion. That is, the higher the naloxone dose, the higher the effect size for patients. Effect sizes for naloxone placebo group patients were near zero, as expected, since true change in the constipation condition is not expected for the placebo patients.

The SEM is a characteristic of the measure for the entire group of patients. The SEM for all patients was 9.01. One half of one SD, based on Visit 3 data, was 11.3. Together these results suggest that score changes below 9 points are not likely clinically meaningful. Score changes of 11.0 points and above may be related to clinically meaningful changes in the constipation condition from the individual patient point of view. It should be noted that treatment may have an important effect on patients even when the mean difference between treatment and control groups is considerably less than the smallest change found clinically meaningful (e.g., Guyatt et al. 1998). The values reported here are estimates to aid in score interpretation. However, based on these two pieces of evidence, score changes of 11 points or greater are likely clinically significant. Interpretation of score differences between 9 and 11 points requires further evaluation. The BFI can detect meaningful change in the constipation condition.

5.0 OVERALL CONCLUSIONS

The BFI is a brief patient rating of constipation. The data reported here support its psychometric properties, necessary information for interpretation of any data based on the BFI. Based on data from this trial, specific BFI score changes can be used as the basis for establishing thresholds for clinically meaningful change in constipation.

6.0 REFERENCES

Cohen J. Statistical power analyses for the behavioral sciences (2nd Ed.) Hillsdale NJ: Erlbaum, 1988.

Deyo RA, Dieher P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evalution. Cont Clin Trials 1991;12:142S-158S.

Drossman DA, Corazziari E, Talley NJ, Thompson WG, Whitehead WE, Rome II Multinational Working Teams. *Rome II: The Functional Gastrointestinal Disorders*. 2nd ed. McLean, VA: Degnon Associates; 2000.

Guyatt GH, Juniper E, Walter S, Griffith L, Goldstein R. Interpreting treatment effects in randomised trials. Br Med J 1998:316(7132);690-693.

Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40(2):171-178.

Hays RD, Anderson RT, Revicki DA. Assessing reliability and validity of measurement in clinical trials. In: Staquet MJ, Hays RD, Fayers PM, eds. *Quality of Life Assessment in Clinical Trials: Methods and Practice*. Oxford: Oxford University Press; 1998.

Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(3 Suppl):S178-89.

Leidy NK, Revicki DA, Geneste B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. *Value in Health*. 1999;2(2):113-127.

Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation. Med Care 2003;41:582-592.

Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York: McGraw-Hill; 1994.

Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *QOL Research*. 2000;9(8):887-900.

Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52:861-873.

7.0 TABLES

Table 1. Demographic Characteristics at Visit 2

Baseline Characteristics	
Age, years	
N	202
Mean (SD)	56.3 (13.1)
Range	27 – 86
Gender, N (%)	
Female	127 (62.9%)
Male	75 (37.1%)
Race, N (%)	
White	202 (100.0%)

Table 2. Clinical Characteristics at Visit 3[a]

Clinical Characteristics	N	Mean (SD)	Median	Range
Daily pain intensity[b]	202	38.3 (16.3)	37.8	0.0 - 80.7
Stool frequency[c]	202	1.0 (0.5)	0.9	0.1 - 4.0
Stool consistency[d]	202	2.4 (0.7)	2.5	1.0 - 4.0
Laxative intake (mean # of days)[e]	191	6.0 (1.9)	7.0	0.0 - 7.0

[a]Average from daily diary over previous seven days.
[b]Average daily pain intensity on a scale of 0 (no pain) to 100 (worst pain imaginable).
[c]Average number of bowel evacuations per day.
[d]Average patient rating on a 4-point response scale (hard=1, solid=2, semi-solid=3, and diarrheal=4).
[e]Average number of days over the previous seven days.

Table 3. Diarrhea as Reason for Discontinuation

N	% of all enrolled subjects
11	5.4%

Table 4. Constipation Item Analysis, Visit 3 and Visit 5

Visit 3	N	Mean (SD)	Range	Median	% at Floor	% at Ceiling	% Missing
Ease of defecation	202	53.4 (24.5)	0 - 100	50.0	3.5%	4.0%	0.0%
Feeling of incomplete bowel evacuation	202	41.3 (28.8)	0 - 100	40.0	18.3%	2.5%	0.0%
Judgment regarding constipation	202	52.5 (24.3)	0 - 100	50.0	5.0%	3.5%	0.0%
Total Score	202	49.1 (22.6)	0 - 100	50.0	2.0%	2.0%	0.0%
Visit 5	**N**	**Mean (SD)**	**Range**	**Median**	**% at Floor**	**% at Ceiling**	**% Missing**
Ease of defecation	169	39.5 (27.0)	0 - 100	40.0	14.2%	1.2%	0.0%
Feeling of incomplete bowel evacuation	169	30.3 (27.0)	0 - 100	30.0	26.6%	1.2%	0.0%
Judgment regarding constipation	169	38.1 (28.6)	0 - 100	40.0	17.8%	1.8%	0.0%
Total Score	169	36.0 (25.0)	0 - 100	33.3	11.8%	0.6%	0.0%

Table 5. Inter-Item Correlations and Internal Consistency Reliability: Constipation Items, Visit 2

Item/Total Score	N	Ease of defecation	Feeling of incomplete bowel evacuation	Judgment regarding constipation	Total Score	Cronbach's alpha[a]
Ease of defecation	202	1.00	0.59[3]	0.86[3]	0.91[3]	0.75
Feeling of incomplete bowel evacuation	202		1.00	0.60[3]	0.84[3]	0.92
Judgment regarding constipation	202			1.00	0.91[3]	0.74
Total Score	202				1.00	0.87

[a]Total score alpha and alpha with item deleted for individual items.
[1]p<0.05 [2]p<0.01 [3]p<0.001

Table 6a. Reproducibility: Constipation Items, Visit 5 and Visit 6 Data for Patients Randomized to Naloxone Placebo Group

	N	Visit 5 Mean (SD)	Visit 6 Mean (SD)	Difference	ICC[a]	Pearson's correlation (*r*)[b]
Ease of defecation	43	49.5 (24.2)	52.7 (27.5)	3.1	0.53	0.53[3]
Feeling of incomplete bowel evacuation	43	37.0 (29.3)	41.7 (27.5)	4.8	0.72	0.72[3]
Judgment regarding constipation	44	49.9 (26.2)	53.5 (26.4)	3.6	0.63	0.63[3]
Total score	44	45.6 (22.5)	50.0 (24.4)	4.4	0.64	0.65[3]

[a]Intraclass correlation coefficient.
[b]Pearson product-moment correlations.
[1]p<0.05 [2]p<0.01 [3]p<0.001

Table 6b. Reproducibility: Constipation Items, Visit 5 and Visit 6 Data for Stable[a] Patients Randomized to Naloxone Placebo Group

	N	Visit 5 Mean (SD)	Visit 6 Mean (SD)	Difference	ICC[b]	Pearson's correlation (*r*)[c]
Ease of defecation	12	48.3 (18.0)	50.8 (23.5)	2.5	0.29	0.28
Feeling of incomplete bowel evacuation	12	31.7 (22.1)	40.4 (21.6)	8.8[1]	0.82	0.88[3]
Judgment regarding constipation	12	49.2 (17.3)	50.4 (24.7)	1.3	0.14	0.14
Total score	12	43.1 (15.0)	47.2 (21.5)	4.2	0.49	0.51

[a]Stability defined as no change in stool frequency.
[b]Intraclass correlation coefficient.
[c]Pearson product-moment correlations.
[1]p<0.05 [2]p<0.01 [3]p<0.001

Table 7. Concurrent Validity: Correlation[a] between Constipation Items and Clinical Data, Visit 5

Item/Total Score	N	Stool frequency	Stool consistency	# Days on laxative	Global tolerability assessment
					Patient rated
Ease of defecation	169	-0.28[3]	-0.23[2]	0.42[3]	0.23[2]
Feeling of incomplete bowel evacuation	169	-0.23[2]	-0.26[3]	0.33[3]	0.34[3]
Judgment regarding constipation	169	-0.28[3]	-0.20[1]	0.44[3]	0.29[3]
Total Score	169	-0.29[3]	-0.24[2]	0.45[3]	0.31[3]

[a]Spearman's rank correlations.
[1]p<0.05 [2]p<0.01 [3]p<0.001

Table 8. Concurrent Validity: Constipation Item Score Comparisons, Data from Visit 5[a]

Item/Total Score	Patients who Prefer Maintenance Therapy Mean (SD) N = 89	Patients who prefer Titration Phase Mean (SD) N = 40	P Value
Ease of defecation	33.4 (23.8)	52.4 (30.9)	0.0011
Feeling of incomplete bowel evacuation	26.3 (23.8)	39.0 (33.3)	0.0339
Judgment regarding constipation	31.5 (25.5)	53.0 (31.0)	0.0003
Total Score	30.4 (22.4)	48.1 (28.7)	0.0010

[a]T-test comparison

Table 9. Discriminant Validity: Constipation Severity based on Stool Consistency Rating in Diary (Analysis of Variance[a]), Visit 5

Item/Score	Mild (N = 94) Mean (SD)	Moderate (N = 53) Mean (SD)	Severe (N = 26) Mean (SD)	Overall F Value	Pairwise Comparisons[b]
Ease of defecation	36.6 (28.1)	36.0 (22.3)	58.8 (24.6)	5.88[2]	B[2], C[2]
Feeling of incomplete bowel evacuation	27.2 (26.9)	24.9 (22.0)	53.8 (26.3)	7.00[2]	B[2], C[2]
Judgment regarding constipation	35.3 (30.1)	34.0 (22.4)	57.9 (27.1)	5.02[2]	B[1], C[1]
Total Score	33.0 (25.3)	31.6 (19.5)	56.8 (24.9)	7.40[3]	B[2], C[2]

[a] ANOVA model includes constipation severity as the independent variable and Visit 5 constipation item and total scores as dependent variables; Visit 3 constipation item scores included as covariates.
[b] P values are: [1]<0.05, [2]<0.01, [3]<0.001. Comparisons are: A: Mild vs. Moderate; B: Mild vs. Severe; C: Moderate vs. Severe.

Table 10. Responsiveness of Constipation Items, Visit 3 and Visit 5, by Treatment Group

Item / Total Score	Naloxone Placebo		10 mg Naloxone		20 mg Naloxone		40 mg Naloxone		Total Group SEM[c] (N = 169)
	Effect Size (Method 1)[a] (N = 45)	Effect Size (Method 2)[b] (N = 45)	Effect Size (Method 1)[a] (N = 41)	Effect Size (Method 2)[b] (N = 41)	Effect Size (Method 1)[a] (N = 43)	Effect Size (Method 2)[b] (N = 43)	Effect Size (Method 1)[a] (N = 40)	Effect Size (Method 2)[b] (N = 40)	
Ease of defecation	0.09	0.09	0.60	0.53	0.87	0.87	1.05	1.32	--
Feeling of incomplete bowel evacuation	0.14	0.15	0.33	0.33	0.66	0.59	0.64	0.68	--
Judgment regarding constipation	0.07	0.09	0.68	0.58	0.83	0.81	1.18	1.39	--
Total Score	0.12	0.14	0.60	0.54	0.93	0.85	1.07	1.22	9.01

[a] Method 1 for computing effect size: (mean Visit 3 score – mean Visit 5 score) / standard deviation of Visit 3 scores of all patients
[b] Method 2 for computing effect size: (mean Visit 3 score – mean Visit 5 score) / standard deviation of Visit 3 scores among stable patients only. Stable patients are defined as those who have less than or equal to 25% decrease on the judgment of constipation item from Visit 3 to Visit 5. Ns for stable subjects range from 12 to 31 for the different treatment groups.
[c] Method for computing SEM: standard deviation multiplied by the square root of (1-Cronbach's alpha). SEM cannot be computed for single items.